DOCUMENT RESUME

ED 411 262 TM 027 207

AUTHOR Sun, Anji; Valiga, Michael J.
TITLE Assessing Reliability of Student Ratings of Advisor: A Comparison of Univariate and Multivariate Generalizability Approaches.
PUB DATE 1997-03-00
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Advising; *Colleges; *Faculty Advisers; *Generalizability Theory; Higher Education; Mathematical Models; Multivariate Analysis; *Reliability; *Student Evaluation of Teacher Performance
IDENTIFIERS Act Survey of Academic Advising

ABSTRACT

In this study, the reliability of the American College Testing (ACT) Program's "Survey of Academic Advising" (SAA) was examined using both univariate and multivariate generalizability theory approaches. The primary purpose of the study was to compare the results of three generalizability theory models (a random univariate model, a mixed univariate model, and a multivariate model) and examine their utility for assessing the reliability of the data. Data were from SAA history files for 10 postsecondary institutions, with 10 advisors from each institution, and 10 students for each advisor. Results of the study demonstrate empirically that, when there is only one item for each level of the fixed facet in the multivariate model, the multivariate G coefficients are identical to the G coefficients produced by the mixed univariate model in which the item is a fixed facet. Although the multivariate and mixed univariate models produced higher G coefficients than did the random univariate model, the choice of using a univariate or multivariate approach should depend on the nature of the items and scales in the instrument and the assumptions underlying the universe of generalization. However, since the multivariate model (with one item per level of the fixed facet) yields the same G coefficient as the mixed univariate model (with item facet fixed) and since statistical procedures for the univariate model are simpler and more available, the mixed univariate model could be used to determine the desired level of generalizability/ reliability of the measurement data. (Contains 4 tables and 11 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 411 262

Assessing Reliability of Student Ratings of Advisor:

A Comparison of Univariate and Multivariate Generalizability Approaches

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
Anji Sun

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

By

Anji Sun

Michael J. Valiga

ACT

PO Box 168

Iowa City, IA 52243

A paper prepared for the Annual Meeting of American Educational Research Association

Chicago

March, 1997

(Please do not cite or quote without permission of the authors)

TM027207

BEST COPY AVAILABLE

Abstract

In this study, the reliability of the ACT *Survey of Academic Advising* (SAA) was examined using both univariate and multivariate generalizability theory approaches. The primary purpose of the study was to compare the results of three generalizability theory models, i.e., a random univariate model, a mixed univariate model, and a multivariate model, and examine their utility for assessing the reliability of the data. Results of the study empirically demonstrate that, when there is only one item for each level of the fixed facet in the multivariate model, the multivariate G coefficients are identical to the G coefficients produced by the mixed univariate model in which item is a fixed facet. Although the multivariate and mixed univariate models produced higher G coefficients than did the random univariate model, the choice of using a univariate or multivariate approach should depend on the nature of the items and scales in the instrument and the assumptions underlying the universe of generalization. However, since the multivariate model (with one item per level of the fixed facet) yields same G coefficient as the mixed univariate model (with item facet fixed), and since statistical procedures for the univariate model are simpler and more available, the mixed univariate model could be used to determine the desired level of generalizability/reliability of the measurement data.

The reliability of student ratings of faculty performance has been typically investigated using procedures arising from two theoretical frameworks. The first type of procedure yielding such reliability indices as coefficient alpha, KR-20, and interrater correlations, was developed within the framework of classical test theory. These classical procedures have been criticized for failing to deal with multiple sources of random error and for disparities between the type of reliability estimated and the decisions that the data are to inform (Marsh, 1982). The second type of procedure was developed within the framework of generalizability theory (Cronbach, Gleser, Nanda, and Rajaratnam, 1972). Because generalizability theory (hereafter referred to as G theory) emphasizes that measurement error is relative and has multiple sources, it is conceptually more meaningful, statistically more powerful, and practically more effective than the classical test theory. For these reasons, generalizability theory has gained growing attention as a framework on which to base studies of the reliability of student ratings of faculty/instructor/advisor performance (Aubretch, 1981; Gillmore, Kane, & Naccarato, 1978; Gillmore, 1983).

Within the G theory framework, a number of possible approaches may be utilized to estimate the generalizability/reliability of the measures in any given study. The decisions as to which approach is more appropriate depends on the object of measurement, the nature of the items and scales in the measurement instrument and the universe of generalization corresponding to the intended uses of the data.

In this study, two generalizability theory approaches, one being univariate and the other multivariate, were compared to determine which was more appropriate for estimating

the reliability of student ratings of advisors collected using the ACT *Survey of Academic Advising* (SAA). The fundamental distinction between the univariate and multivariate approaches lies in the number of universe scores associated with the objects of measurement. In the univariate approach, each object of measurement is associated with only one universe score, which can be viewed as "a mean score for an object of measurement over all conditions in the universe of generalization." (Brennan, 1992) However, in the multivariate approach, "each object of measurement has multiple universe scores, with each such universe score associated with a specific level of a fixed facet." (Brennan, 1992)

The advantage of the multivariate over the univariate generalizability approach for examining the reliability of the measurement data is that the multivariate approach not only takes into account of the variance component, but also the covariance components, for each person's universe score. In addition, this approach takes into account the correlated error when estimating error variances. Moreover, multivariate analysis provides the researcher with more useful information about the data, including disattenuated correlations for pairs of measurements and each measure's relative contribution to the universe score variance and error variance, that cannot be obtained using the univariate models (Brennan, 1992; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Joe & Woodword, 1976; Shavelson & Webb, 1981; 1991).

Smith and Kane (1981) provided an in-depth discussion about whether to choose the univariate approach or the multivariate approach for reliability analysis. They stated that since "a set of scores could be considered multivariate with respect to a set of fairly narrowly defined constructs, and univariate (i.e., multiple observations of the same construct) with

respect to a single broadly defined construct", the choice of analytical method should depend on how the data are to be used for making decisions or inferences. Specifically, if the scores are intended to be interpreted as multiple observations of the same construct, "then differences in the conditions for arriving at this scores should be considered conditions of a facet and a univariate procedure is appropriate." However, if the scores are to be interpreted as a pattern or profile, each member of which implies a somewhat different construct, then a multivariate or profile analysis would be appropriate.

In the case of the SAA data considered in this study, the primary purpose was to differentiate among the advisors to facilitate personnel decisions, such as salary adjustment, promotion, or tenure, and to provide information to advisors for personal improvement. Another equally important purpose of the SAA study is to evaluate the various academic advising services provided by the institution. For these two purposes, developers of the SAA instrument constructed 18 items, each of which describes a specific, unique area of academic advising. Students who provided data for the study were asked to indicate their level of satisfaction with their advisor's assistance on each of the 18 items, using a 5-point Likert scale.

One issue involved in deciding whether to use the univariate or the multivariate generalizability approach for the SAA study centers on how the item facet should be viewed and treated in the analysis. If the 18 items are viewed as 18 distinctive areas of academic advising, then the assumptions of the multivariate approach seems appropriate and the items should be treated as finite levels of the fixed facet of academic advising. On the other hand, if the 18 items are viewed as 18 possible descriptors for one general area, i.e., academic

advising, then they should be treated as a random or fixed facet. In this situation, the univariate approach would be more appropriate.

As described before, the SAA items represent a broad array of aspects of academic advising. Each item can be viewed as a content category of academic advising with the 18 items basically exhausting the content of academic advising. For this reason, the items are probably best viewed as fixed sample from the advising universe, so the univariate approach does not seem to fit. However, there is a problem in applying the multivariate approach. That is, since there is only one item in each of the 18 advising areas, there is no way to isolate and estimate the random error introduced by the item facet. Also, the interaction between advisor and item and the interaction between student and item cannot be estimated. Consequently, the error variances will likely be underestimated and the estimated generalizability coefficients will be somewhat inflated.

Although the multivariate generalizability model seems to be most appropriate for assessing the reliability of the SAA data, it is informative to compare it with the univariate approach. For this reason, a random univariate model, a mixed univariate model, and a multivariate model were considered in this study. Results of the three models can shed light on the advantages and disadvantages in using each of the approaches for assessing the reliability of the SAA or similar data.

Method

Data

Data for the study were obtained from the SAA history files maintained by American College Testing. Ten of the 118 postsecondary institutions in the United States that had

administered the SAA between 1992 to 1994 were randomly selected for the study. Within each institution, 10 advisors were randomly selected from all of the advisors who had advised more than 20 students within the past three years. For each of the 10 advisors, 10 students were randomly selected from all of the students who had received the advisor's assistance. The advisors were rated, using a 5-point Likert scale, in term of students' satisfaction with the advisor's assistance in 18 different areas. For analysis, all responses were coded as integer values, with *Very Dissatisfied* coded as "1" and *Very Satisfied* coded as "5". Because balanced data were required for the procedure, missing responses were replaced with group means.

Models

The *G* study design for estimating the variance components using the univariate approach can be represented as (s:a:c) x i, where "s" is the student, "a" is the advisor, "c" is the institution, and "i" is the item facet. Estimates of seven variance components were produced by the univariate *G* study analysis. For the multivariate model, since there is only one item in each of the 18 advising areas, the variance-covariance components for items cannot be estimated. Because the variance component for the institution effect was observed to be quite small, there was little need to include the institution facet in our model. Therefore, the variance and covariance components for the advisor and student facets were first calculated for each institution separately, then pooled together by averaging over the 10 institutions. The multivariate *G* study yielded two 18 x 18, variance-covariance component matrices: one for the advisor effect and the other for the student nested within advisor effect.

Three *D* study models were used in this study: the (S:a) x I random univariate model (both facets random), the (S:a) x I mixed univariate model (student facet random but item facet fixed), and the S:a multivariate model in which items is a hidden facet. Error variances and generalizability coefficients for the composite SAA rating scores with differing number of students were estimated using both the univariate and multivariate *D* study designs. The formula for the G coefficient for the random univariate model is

$$\varepsilon\hat{\rho}_a^2 = \frac{\hat{\sigma}^2_{a:c}}{\hat{\sigma}^2_{a:c} + \frac{\hat{\sigma}^2_{s:a:c}}{n'_{s:a:c}} + \frac{\hat{\sigma}^2_{ai:c}}{n'_i} + \frac{\hat{\sigma}^2_{si:a:c}}{n'_{s:a:c}n'_i}}$$

The formula for G coefficient for the mixed univariate model is

$$\varepsilon\hat{\rho}_a^2 = \frac{\hat{\sigma}^2_{a:c} + \frac{\hat{\sigma}^2_{ai:c}}{n'_i}}{\hat{\sigma}^2_{a:c} + \frac{\hat{\sigma}^2_{ai:c}}{n'_i} + \frac{\hat{\sigma}^2_{s:a:c}}{n'_{s:a:c}} + \frac{\hat{\sigma}^2_{si:a:c}}{n'_{s:a:c}n'_i}}$$

The formula for G coefficient for the multivariate model is

$${}_w\varepsilon\rho^2 = \sigma^2({}_w\mu_a)/[\sigma^2({}_w\mu_a) + \sigma^2({}_w\delta)]$$

Where

$$\sigma^2({}_w\mu_a) = \Sigma\, w^2_v\sigma^2({}_va) + \Sigma\,\Sigma\, 2\, w_v w_{v'}\sigma({}_va_{v'}a);$$

$$\sigma^2({}_w\delta) = \Sigma\, w^2_v\sigma^2({}_v\delta_{S:a}) + \Sigma\,\Sigma\, 2\, w_v w_{v'}\sigma({}_v\delta_{S:a_{v'}}\delta_{S:a});$$

$\sigma^2 ({}_v\delta_{S:a}) = \sigma^2 ({}_vS{:}a)/n'_{v:a}$; and

$\sigma ({}_v\delta_{S:a,v'}\delta_{S:a}) = \sigma ({}_vS{:}a_{v'}S{:}a)/n'_{v:a}$.

Results and Discussions

Univariate Generalizability Model

Table 1 presents the estimated variance components for the (s:a:c) x i *G* study design. To show the relative contribution to the total variance by each effect, the table also presents the proportion of total variability accounted for by each of the variance components (Shavelson & Webb, 1991). As seen in Table 1, the largest source of variation in the model was the si:a:c residual effect accounting for 41% of the total variance. The second largest variance component was the student effect, which accounted for 29% of the total variance, an indication that individual differences among students played an important part in how satisfied they were about their advisor's assistance in academic advising. The third largest variance component was for advising, which accounted for 15% of total variance, suggesting that a considerable proportion of the variability in students' satisfaction ratings was attributable to the advisors. The variance component for interaction between advisor and item accounted for 9% of the total variance, which seems to suggest that advisors were rated differently across different items. The item facet did not seem to contribute to the total variability very much, since it only accounted for 4% of the total variance. The variance components for college and college by item interaction were close to zero.

Results of the univariate *D* study models are presented in Table 4. As seen in Table 4, the SAA ratings are quite reliable, regardless of models, if a minimum of 10 students are sampled for each advisor. The generalizability coefficients using 10 students per advisor are

.81 for the random model and .84 for the mixed model. It should be pointed out that, although the mixed model yielded higher generalizability coefficients, the universe of generalization is limited to only the 18 items. In other words, the results of the study cannot be generalized beyond the 18 items used in the *G* study.

The multivariate generalizability model

Variance-covariance components for advisor and for student in the s:a multivariate *G* study model are presented in Tables 2 and 3. A comparison of the sizes of the variance components (diagonal elements of the variance-covariance component matrix) on both matrices suggests that, although the student facet was the largest source of total variance, the difference among advisors contribute substantially to the variability of the SAA ratings in most of the areas. More importantly, the covariance components (off diagonal elements of the variance-covariance component matrix) for advisor were substantial, reflecting the underlying correlations among the 18 areas of academic advising. On the one hand, advisors' assistance seen by his/her students as satisfactory on one area is likely to be seen as satisfactory on other areas. On the other hand, this could be an indication of a response set or a halo effect, as suggested by previous studies of student ratings, i.e., students' ratings on particular and apparently separate characteristics of a professor are affected by their general, overall characteristics and attitudes toward the professor (Cruse, 1987).

A comparison of the univariate and multivariate models

Results of the univariate *D* studies were compared with those of multivariate *D* studies. The variance component for the advisor effect was .0846 for the random univariate model, .08741 for the mixed univariate model, and .08741 for the multivariate model. With a

sample size of 10 students per advisor, the univariate G coefficient was .81 for the random model and .83 for the mixed model. The multivariate G coefficient was .83, identical to that of the mixed model.

The results of the study demonstrate that when there is only one item within each level of a fixed facet, the multivariate generalizability coefficient is identical to that of the mixed univariate model in which item is a fixed facet. The identical G coefficients may result from the fact that, in the mixed univariate model, the variance component for universe score equals the variance component for advisor plus the variance component for interaction between advisor and item corrected by the item sample size, which equals the variance-covariance component for advisor in the multivariate model. Similarly, the variance component for relative error score in the mixed model equals the variance component for student plus the variance component for interaction between student and item corrected by the item sample size, which equals the variance-covariance component for relative error score in the multivariate model. Consequently, the generalizability coefficients of the two models, defined as the ratio of the variance component for advisor and the observed variance in the univariate model and the variance-covariance component for advisor and the observed variance-covariance in the multivariate model, are the same. Further research/mathematical modeling is needed to demonstrate algebraically the relationship between these two models.

It is important to emphasize that the size of G coefficient should not be a factor in determining whether to choose the univariate approach or the multivariate approach. The choice must depend on the nature of the instrument and considerations for the universe of generalization. However, since the multivariate model (with one item per level of the fixed

facet) yields same G coefficient as the mixed univariate model (with item facet fixed), and since statistical procedures for the univariate model are simpler and more available, the mixed univariate model could be used to determine the desired level of generalizability/reliability of the measurement data.

References

Aubrecht, J. D. (1981). Reliability, Validity, and Generalizability of Student Ratings of Instruction. IDEA Paper, No.6, Center for Faculty Evaluation & Development in Higher Education. Kansas State University, Manhattan.

Brennan, R. L. (1992). *Elements of Generalizability Theory* (rev.ed.). Iowa City, IA: American College Testing.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cruse, D. B. (1987). Student evaluations and the university professor: Caveat Professor. Higher Education, 16, 723-737.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizability of student ratings of instruction: Estimation of the teacher and course components. Journal of Educational Measurement, 15(1), 1-13.

Gillmore, G. M. (1983). Generalizability theory: Applications to program evaluation. *New direction for testing and measurement: Generalizability theory, 18*, 3-16.

Joe, G. W. & Woodward, J. A. (1976). Some development in multivariate generalizability. Psychometrika, 41(2), 205-217.

March, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. Journal of educational psychology, 74, 264-279.

Shavelson, R. J., & Webb, N. M. (1991). Generalizability theory: A primer. Newbury Park, CA: Sage.

Smith, P. L. & Kane, M. (1981). Multivariate generalizability: An application to student course ratings. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA.

Webb, N. M. & Shavelson, R. J. (1981). Multivariate generalizability of general educational development ratings. Journal of educational measurement, 18(1), 13-22.

Table 1

Estimated Variance Components for the Univariate Generalizability Study of the SAA Data

Source of Variability	[a]n	VC[b]	% Total Var[c]	SE
College (c)	10	.00235	.040	.00532
Advisor:college (a:c)	10	.08460	15.415	.01018
Student:advisor:college (s:a:c)	10	.16035	29.218	.00669
Item (i)	18	.02414	4.399	.00885
College x item (ci)		.00245	.446	.00105
Advisor:college x item (ai:c)		.05063	9.225	.00223
Student:advisor:college x item (si:a:c)		.22429	40.868	.00219

[a]n=sample size in *G* study estimation of variance components. [b]Estimated variance component. [c]Percent of total variance.

Table 2. Estimated Variance and Covariance Components for Advisor on 18 Academic Advising Areas

Area	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**.082**																	
2	.087	**.111**																
3	.067	.087	**.099**															
4	.081	.099	.094	**.150**														
5	.078	.093	.072	.083	**.144**													
6	.094	.114	.087	.107	.097	**.146**												
7	.059	.076	.068	.098	.075	.082	**.110**											
8	.072	.098	.088	.090	.102	.090	.095	**.146**										
9	.041	.055	.054	.064	.075	.046	.074	.077	**.119**									
10	.072	.092	.068	.079	.095	.085	.085	.083	.077	**.116**								
11	.091	.102	.062	.084	.105	.092	.075	.100	.059	.101	**.134**							
12	.080	.097	.095	.096	.100	.098	.089	.109	.073	.101	.102	**.141**						
13	.072	.102	.095	.098	.107	.086	.088	.110	.096	.090	.095	.113	**.191**					
14	.060	.070	.062	.068	.078	.061	.080	.094	.065	.083	.074	.079	.097	**.115**				
15	.066	.068	.070	.084	.083	.083	.063	.084	.051	.072	.087	.084	.114	.087	**.127**			
16	.068	.086	.074	.086	.106	.107	.085	.100	.078	.094	.096	.111	.126	.079	.100	**.160**		
17	.057	.056	.053	.086	.090	.085	.067	.098	.072	.057	.082	.085	.103	.075	.100	.132	**.189**	
18	.055	.078	.068	.087	.079	.082	.085	.094	.076	.069	.072	.098	.130	.087	.087	.117	.105	**.155**

Note. Bold elements are variance components, non-bold elements are covariance components.

Table 3. Estimated Variance and Covariance Components for Student on 18 Academic Advising Areas

Area	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**.465**																	
2	.263	**.645**																
3	.204	.277	**.464**															
4	.128	.117	.133	**.312**														
5	.157	.181	.166	.144	**.367**													
6	.237	.251	.206	.142	.157	**.514**												
7	.155	.141	.130	.156	.123	.160	**.341**											
8	.176	.171	.186	.163	.157	.201	.231	**.362**										
9	.113	.102	.116	.143	.125	.139	.180	.178	**.272**									
10	.187	.191	.162	.157	.175	.213	.191	.212	.157	**.430**								
11	.190	.184	.164	.159	.166	.209	.173	.210	.155	.287	**.394**							
12	.214	.193	.167	.143	.145	.209	.205	.202	.161	.203	.210	**.414**						
13	.109	.130	.106	.135	.110	.160	.172	.158	.150	.160	.170	.156	**.366**					
14	.132	.122	.104	.145	.104	.141	.172	.160	.153	.155	.167	.180	.194	**.334**				
15	.146	.140	.126	.161	.132	.167	.171	.171	.147	.179	.191	.191	.177	.195	**.334**			
16	.159	.146	.122	.143	.132	.176	.139	.151	.120	.172	.189	.167	.163	.170	.208	**.321**		
17	.119	.102	.101	.117	.123	.120	.134	.128	.118	.121	.130	.148	.138	.137	.138	.158	**.273**	
18	.149	.123	.130	.132	.130	.161	.156	.157	.134	.165	.166	.196	.165	.152	.171	.155	.150	**.316**

Note. Bold elements are variance components, non-bold elements are covariance components.

BEST COPY AVAILABLE

Table 4

Estimated Generalizability Coefficients for the SAA Data with Various Student Sample Sizes

Student Sample Size	Random Model G Coefficient	SE	Mixed Model G Coefficient	SE	Multivariate Model G-coefficient	SE
5	.693	.193	.717	.186	.717	.186
10	.808	.142	.835	.131	.835	.131
15	.855	.120	.884	.107	.884	.107
20	.880	.108	.910	.093	.910	.093
25	.897	.099	.927	.083	.927	.083
30	.908	.093	.938	.076	.938	.076